

Análisis Inteligente de Datos: Segundo Parcial

Claudio Sebastián Castillo

16 de mayo de 2022

Contents

1	EDA	3
1.0.1	structure	3
1.0.2	Summary	3
1.0.3	Control NAs	4
1.0.4	Distribución de datos	4
1.0.5	Grafico Correlaciones	5
1.0.6	Boxplot variables numericas	6
1.0.7	Multigráficos	7
2	ANOVA	8
2.1	Datos	8
2.2	Observaciones por grupo:	8
2.3	Se cumplen los supuestos para su implementación?	8
2.4	Anova	8
2.4.1	fit del modelo	8
2.4.2	coeficientes	8
2.4.3	p-value	8
2.4.4	F-value	8
2.4.5	Plot ANOVA	8
2.4.6	Conclusión	8
2.5	Testear homosedasticidad	8
2.6	Testear normalidad	8
2.7	Testear normalidad analizando residuos	8
2.8	Anova y después: post-hoc	8
2.9	Tukey's Honest Significant Differences (HSD)	8
2.10	Cuando ANOVA no funciona: test de Kruskal-Wallis	8
3	ANOVA_multivariante	9
3.1	Datos	9
3.2	Gafico	9
3.3	Test Anovam	9
3.4	Tamaño del efecto	9
3.5	Cumplimiento requisitos ANOVAM	9
4	Analisis Discriminante Lineal (LDA)	10
4.1	Datos	10
4.1.1	Box por variable	10
4.2	Explorando discriminación por pares de variable	10
4.3	Histograma VariablexGrupo	10
4.4	Contraste de Normalidad Univariante Shapiro-Wilk	10

4.5	Contraste de Normalidad MultiVariante	10
4.5.1	Outliers	10
4.5.2	Test de Royston	10
4.5.3	Test de Henze-Zirkler	10
4.6	Contraste Homosedasticidad	10
4.6.1	Test sobre Matriz de Covarianza	10
4.6.2	Test de Levene	10
4.7	Estimación de parámetros de la función de densidad y cálculo de la función discriminante según aproximación de Fisher via lda()	10
4.8	Evaluación del error en Test Set: Accuracy Table	10
4.9	Precisión del modelo en test set	10
4.10	Error en test set	10
4.11	Validación Cruzada (leave one out)	10
4.12	Visualización de las clasificaciones	10
5	Analisis Discriminante Cuadrático (QDA)>falta de homocedasticidad/outliers LDA	11
5.1	Explorando discriminación por pares de variable	11
5.2	Contraste de Normalidad Univariante Shapiro-Wilk	12
5.3	Contraste de Normalidad MultiVariante	13
5.3.1	Outliers	13
5.3.2	Test de Royston	14
5.3.3	Test de Henze-Zirkler	14
5.4	Contraste de Matriz de Covarianza	14
5.5	Parámetros de la función de densidad función discriminante según aproximación de Fisher via qda()	14
5.6	Evaluación del error: Accuracy Table	15
5.7	Visualización de las clasificaciones	15
6	Analisis Discriminante Cuadrático Robusto (RQDA)	17
7	Máquinas de Soporte Vectorial	18
7.1	Datos	18
7.2	Grafico datos	18
7.3	Busqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel lineal	18
7.3.1	Mejor modelo según hiperparametro	18
7.4	Predicciones del Modelo	18
7.5	Busqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel polynomial	18
7.5.1	Mejor modelo según hiperparametro	18
7.6	Predicciones del Modelo	18
7.7	Busqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel sigmoid	18
7.7.1	Mejor modelo según hiperparametro	18
7.8	Predicciones del Modelo	18
7.9	Busqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel radial	18
7.9.1	Mejor modelo según hiperparametro	18
7.10	Predicciones del Modelo	18
8	Regresión logística (LR)	19
8.1	Datos	19
8.1.1	Boxplot todas las variables	19
8.1.2	Box por grupo	19
8.1.3	Box por variable	19
8.1.4	P valores	19
9	Clustering	20
9.1	Datos	20

9.1.1	Se centran los datos de la matriz con omisión de datos faltantes	20
9.1.2	Se calculan distancias euclídeas	20
9.2	Heatmap de la matriz de distancias entre observaciones del dataset	20
9.3	Cluster No Jerárquico: K-means (x centroides)	21
9.3.1	Selección de k en base al método elbow.	21
9.3.2	Resultado	21
9.3.3	Grafico Cluster con PCA	23
9.4	Cluster No Jerárquico: K-medoids clustering (con centro en observación más representativa) .	24
9.4.1	Selección de k con distancia de Manhattan como medida de similitud	24
9.5	Cluster Jerárquicos	25
9.6	Modelo óptimo considerando distintas matrices de distancias y linkage intercluster	25
9.7	Estudio de la tendencia de clustering	28
10	PCA	29
10.1	Exploración de datos	29
10.2	Análisis PCA	29
10.3	Varianza explicada por las CP	29
10.4	Representaciones Gráficas: BIPLLOT	29
10.5	Proporción de Varianza Explicada	29
10.6	Proporción acumulada en las componentes	29
10.7	FactomineR	29
10.8	Representación	29
10.9	Representación de Variables	29
10.10	Contribuciones de las variables en los ejes	29
10.11	Conclusión	29
11	Sesión	30

1 EDA

1.0.1 structure

```
tibble [77 x 11] (S3: tbl_df/tbl/data.frame)
 $ País      : chr [1:77] "Brasil" "Alemania" "Italia" "Argentina" ...
 $ Mundiales : num [1:77] 20 18 18 16 14 14 14 10 12 11 ...
 $ Puntos    : num [1:77] 227 218 156 140 99 98 96 93 72 61 ...
 $ PJ        : num [1:77] 104 106 83 77 59 62 59 50 51 46 ...
 $ PG        : num [1:77] 70 66 45 42 29 26 28 27 20 16 ...
 $ PE        : num [1:77] 17 20 21 14 12 20 12 12 12 13 ...
 $ PP        : num [1:77] 17 20 17 21 18 16 19 11 19 17 ...
 $ Rendimiento: num [1:77] 0.73 0.69 0.63 0.61 0.56 0.53 0.54 0.62 0.47 0.44 ...
 $ GF        : num [1:77] 221 224 128 131 92 79 106 86 80 74 ...
 $ GC        : num [1:77] 102 121 77 84 66 56 71 48 71 69 ...
 $ Campeón   : num [1:77] 1 1 1 1 1 1 1 0 1 0 ...
```

1.0.2 Summary

País	Mundiales	Puntos	PJ
Length:77	Min. : 1.000	Min. : 0.00	Min. : 1.00
Class :character	1st Qu.: 1.000	1st Qu.: 4.00	1st Qu.: 5.00
Mode :character	Median : 4.000	Median : 14.00	Median : 13.00
	Mean : 5.519	Mean : 30.16	Mean : 21.71
	3rd Qu.: 9.000	3rd Qu.: 40.00	3rd Qu.: 31.00
	Max. :20.000	Max. :227.00	Max. :106.00

PG	PE	PP	Rendimiento
Min. : 0.000	Min. : 0.000	Min. : 1.000	Min. : 0.0000
1st Qu.: 1.000	1st Qu.: 1.000	1st Qu.: 3.000	1st Qu.: 0.2200
Median : 3.000	Median : 3.000	Median : 7.000	Median : 0.3600
Mean : 8.442	Mean : 4.831	Mean : 8.442	Mean : 0.3313
3rd Qu.: 11.000	3rd Qu.: 7.000	3rd Qu.: 14.000	3rd Qu.: 0.4600
Max. : 70.000	Max. : 21.000	Max. : 25.000	Max. : 0.7300

GF	GC	Campeón
Min. : 0.0	Min. : 2.0	Min. : 0.0000
1st Qu.: 4.0	1st Qu.: 10.0	1st Qu.: 0.0000
Median : 13.0	Median : 22.0	Median : 0.0000
Mean : 30.9	Mean : 30.9	Mean : 0.1039
3rd Qu.: 43.0	3rd Qu.: 48.0	3rd Qu.: 0.0000
Max. : 224.0	Max. : 121.0	Max. : 1.0000

1.0.3 Control NAs

```
# A tibble: 1 x 11
  País Mundiales Puntos PJ PG PE PP Rendimiento GF GC Campeón
<int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
1 0 0 0 0 0 0 0 0 0 0 0
```

1.0.4 Distribución de datos

\$coeficiente_variacion

```
# A tibble: 1 x 10
  Mundiales Puntos PJ PG PE PP Rendimiento GF GC Campeón
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 88.3 149. 107. 161. 105. 74.5 55.6 143. 86.0 296.
```

\$sesgo

```
# A tibble: 1 x 10
  Mundiales Puntos PJ PG PE PP Rendimiento GF GC Campeón
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1.13 2.64 1.80 2.79 1.49 0.660 -0.253 2.55 1.16 2.60
```

\$curtosis

```
# A tibble: 1 x 10
  Mundiales Puntos PJ PG PE PP Rendimiento GF GC Campeón
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 3.45 10.5 6.17 11.5 4.69 2.24 2.40 10.4 3.83 7.74
```

\$mad

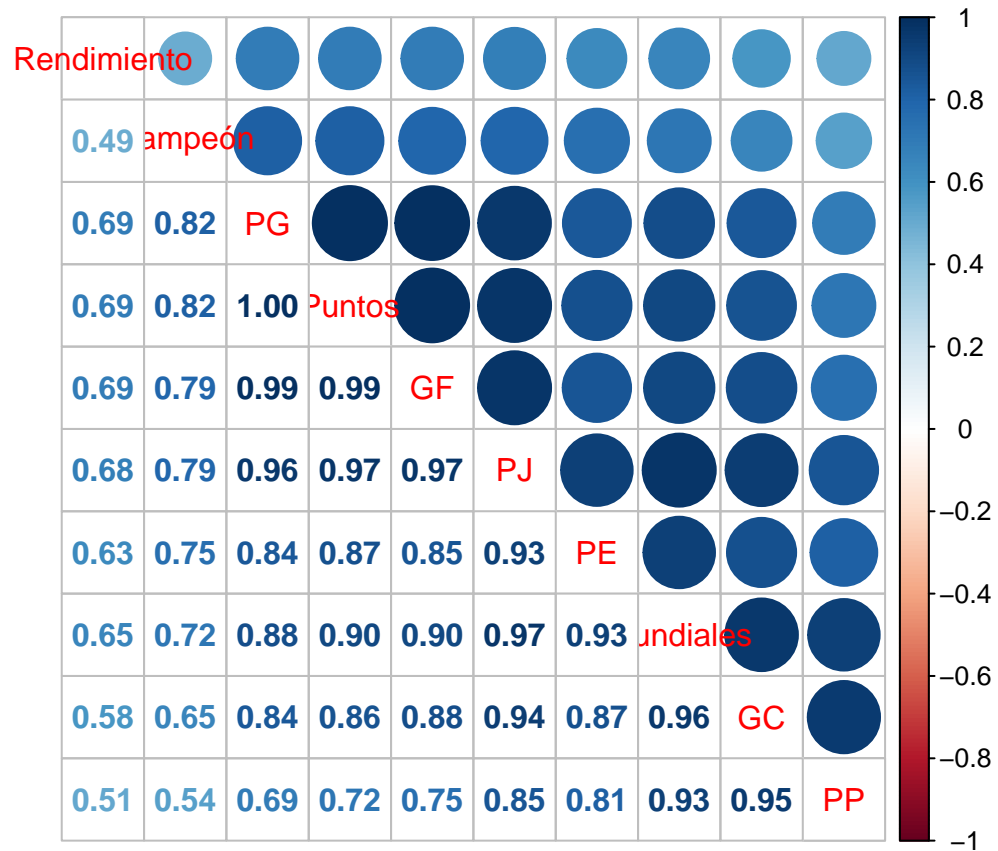
```
# A tibble: 1 x 10
  Mundiales Puntos PJ PG PE PP Rendimiento GF GC Campeón
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 4.45 17.8 14.8 4.45 2.97 5.93 0.193 17.8 23.7 0
```

\$m_correlacion

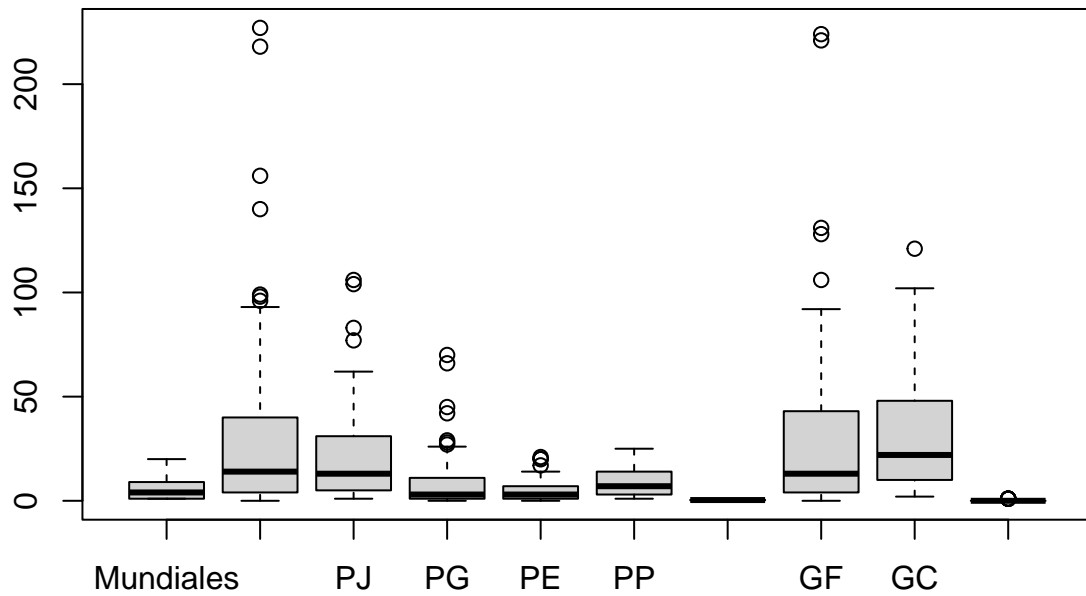
	Mundiales	Puntos	PJ	PG	PE	PP	Rendimiento	GF	GC	Campeón
Mundiales	1.00	0.90	0.97	0.88	0.93	0.93	0.65	0.90	0.96	0.72
Puntos	0.90	1.00	0.97	1.00	0.87	0.72	0.69	0.99	0.86	0.82
PJ	0.97	0.97	1.00	0.96	0.93	0.85	0.68	0.97	0.94	0.79
PG	0.88	1.00	0.96	1.00	0.84	0.69	0.69	0.99	0.84	0.82
PE	0.93	0.87	0.93	0.84	1.00	0.81	0.63	0.85	0.87	0.75

PP	0.93	0.72	0.85	0.69	0.81	1.00	0.51	0.75	0.95	0.54
Rendimiento	0.65	0.69	0.68	0.69	0.63	0.51	1.00	0.69	0.58	0.49
GF	0.90	0.99	0.97	0.99	0.85	0.75	0.69	1.00	0.88	0.79
GC	0.96	0.86	0.94	0.84	0.87	0.95	0.58	0.88	1.00	0.65
Campeón	0.72	0.82	0.79	0.82	0.75	0.54	0.49	0.79	0.65	1.00

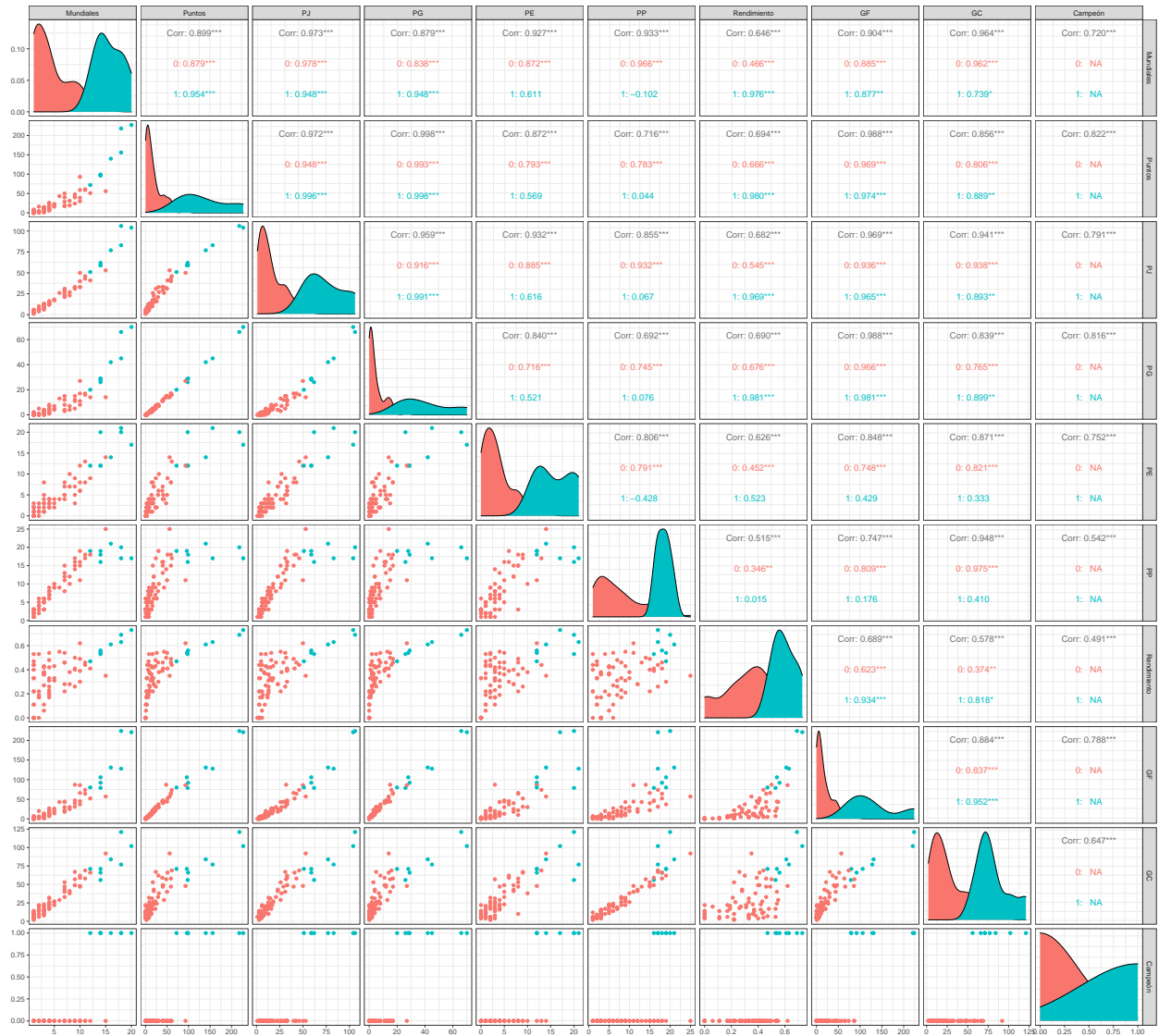
1.0.5 Grafico Correlaciones



1.0.6 Boxplot variables numericas



1.0.7 Multigráficos



2 ANOVA

2.1 Datos

2.2 Observaciones por grupo:

2.3 Se cumplen los supuestos para su implementación?

2.4 Anova

2.4.1 fit del modelo

2.4.2 coeficientes

2.4.3 p-value

2.4.4 F-value

2.4.5 Plot ANOVA

2.4.6 Conclusión

2.5 Testear homosedasticidad

Test de Levene

Test de Bartlett

sensibilidad al supuesto de normalidad

2.6 Testear normalidad

2.7 Testear normalidad analizando residuos

2.8 Anova y después: post-hoc

2.9 Tukey's Honest Significant Differences (HSD)

2.10 Cuando ANOVA no funciona: test de Kruskal-Wallis

3 ANOVA_multivariante

3.1 Datos

3.2 Gafico

3.3 Test Anovam

3.4 Tamaño del efecto

Bajo la siguiente escala del efecto: .01: Small effect size .06: Medium effect size .14 or higher: Large effect size, tenemos:

3.5 Cumplimiento requisitos ANOVAM

4 Analisis Discriminante Lineal (LDA)

4.1 Datos

4.1.1 Box por variable

->

->

->

4.2 Explorando discriminación por pares de variable

4.3 Histograma VariablexGrupo

4.4 Contraste de Normalidad Univariante Shapiro-Wilk

4.5 Contraste de Normalidad MultiVariante

4.5.1 Outliers

4.5.2 Test de Royston

4.5.3 Test de Henze-Zirkler

4.6 Contraste Homosedasticidad

4.6.1 Test sobre Matriz de Covarianza

4.6.2 Test de Levene

4.7 Estimación de parámetros de la función de densidad y cálculo de la función discriminante según aproximación de Fisher via lda()

4.8 Evaluación del error en Test Set: Accuracy Table

4.9 Precisión del modelo en test set

4.10 Error en test set

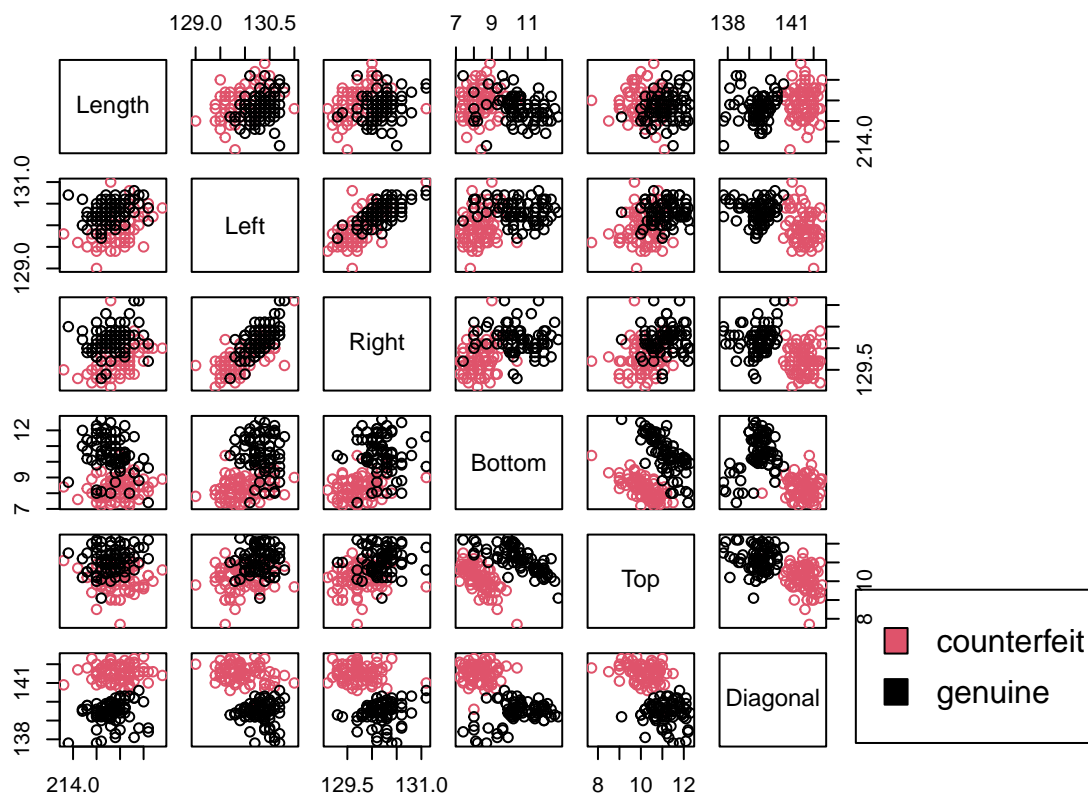
4.11 Validación Cruzada (leave one out)

4.12 Visualización de las clasificaciones

5 Analisis Discriminante Cuadrático (QDA) > falta de homocedasticidad/outliers LDA

```
tibble [200 x 7] (S3: tbl_df/tbl/data.frame)
 $ Status   : Factor w/ 2 levels "counterfeit",...: 2 2 2 2 2 2 2 2 2 2 ...
 $ Length   : num [1:200] 215 215 215 215 215 ...
 $ Left     : num [1:200] 131 130 130 130 130 ...
 $ Right    : num [1:200] 131 130 130 130 130 ...
 $ Bottom   : num [1:200] 9 8.1 8.7 7.5 10.4 9 7.9 7.2 8.2 9.2 ...
 $ Top      : num [1:200] 9.7 9.5 9.6 10.4 7.7 10.1 9.6 10.7 11 10 ...
 $ Diagonal : num [1:200] 141 142 142 142 142 ...
```

5.1 Explorando discriminación por pares de variable



5.2 Contraste de Normalidad Univariante Shapiro-Wilk

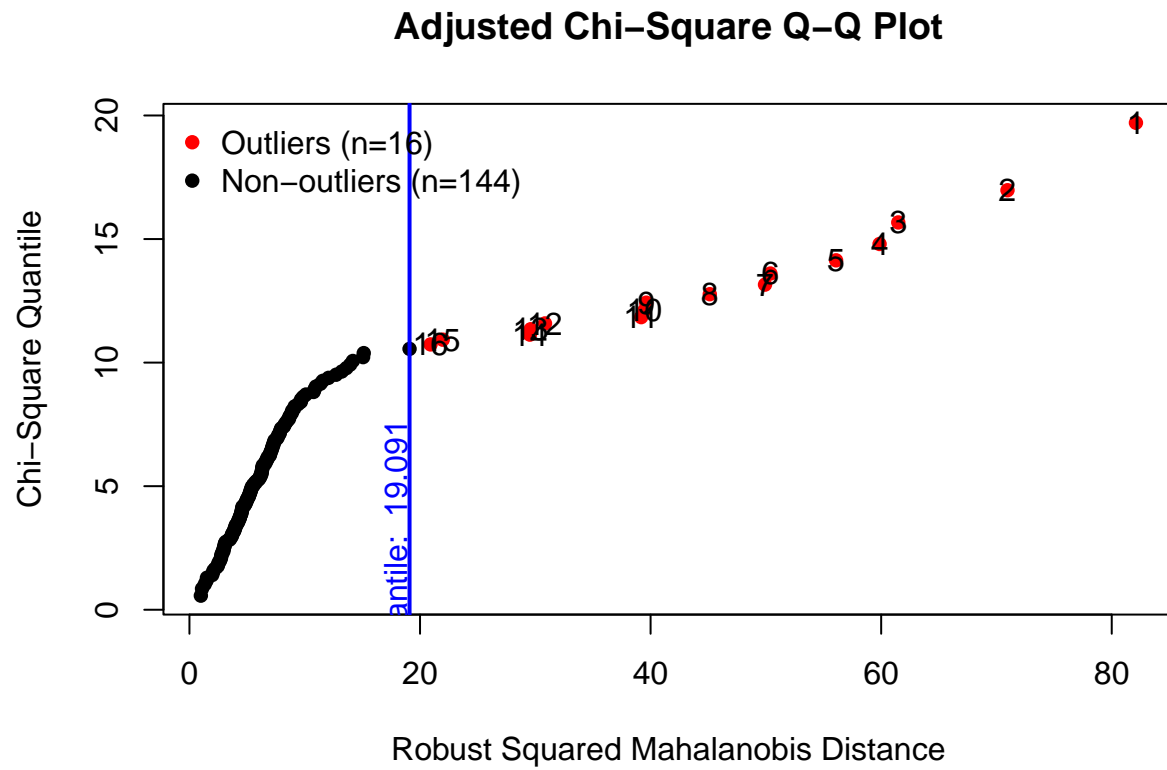
train_tidy[["Status"]]	variable	p_value_Shapiro.test
counterfeit	Length	0.53762
counterfeit	Left	0.11487
counterfeit	Right	0.02634
counterfeit	Bottom	0.03634
counterfeit	Top	0.07829
counterfeit	Diagonal	0.00032
genuine	Length	0.45540
genuine	Left	0.07771
genuine	Right	0.00344
genuine	Bottom	0.08271
genuine	Top	0.02275
genuine	Diagonal	0.00690

[1] "H0 debe rechazarse: hay evidencia de falta de normalidad en los siguientes casos"

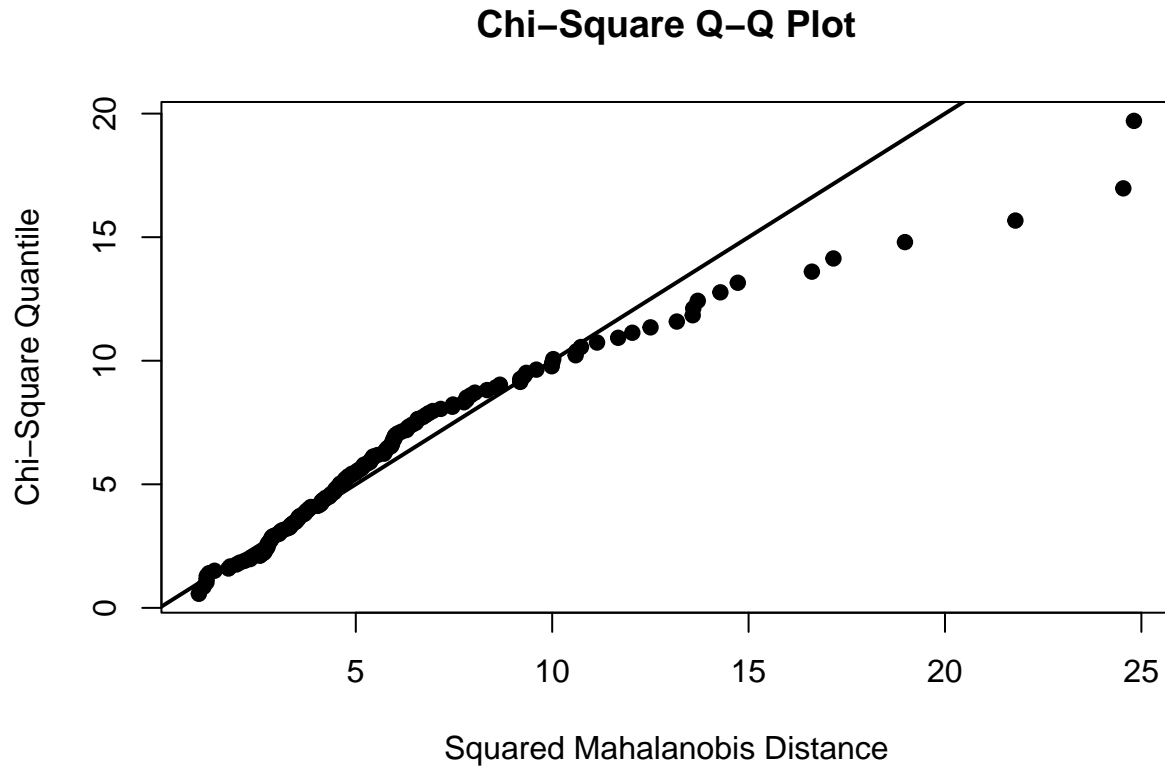
```
# A tibble: 6 x 3
# Groups:   train_tidy[["Status"]] [2]
  `train_tidy[["Status"]]` variable p_value_Shapiro.test
  <fct>                  <fct>          <dbl>
1 counterfeit           Right            0.0263
2 counterfeit           Bottom            0.0363
3 counterfeit           Diagonal           0.00032
4 genuine               Right            0.00344
5 genuine               Top              0.0228
6 genuine               Diagonal           0.0069
```

5.3 Contraste de Normalidad MultiVariante

5.3.1 Outliers



5.3.2 Test de Royston



```
Test      H      p value MVN
1 Royston 50.69247 0.000000001406344 NO
```

```
[1] "H0 debe rechazarse: falta de normalidad multivariante a nivel de significancia 0.05"
```

5.3.3 Test de Henze-Zirkler

```
Test      HZ p value MVN
1 Henze-Zirkler 1.591      0 NO
```

```
[1] "H0 debe rechazarse: falta de normalidad multivariante a nivel de significancia 0.05"
```

5.4 Contraste de Matriz de Covarianza

Box's M-test for Homogeneity of Covariance Matrices

```
data: temp
Chi-Sq (approx.) = 103.03, df = 21, p-value = 0.0000000000008354
```

```
[1] "H0 debe rechazarse: hay evidencia de que la covarianza no es igual en todos los grupos"
```

5.5 Parámetros de la función de densidad función discriminante según aproximación de Fisher via qda()

```
Call:
qda(temp, train[,{
```

```

{
  variable_factor_qda
}
}]]

```

Prior probabilities of groups:

counterfeit	genuine
0.5	0.5

Group means:

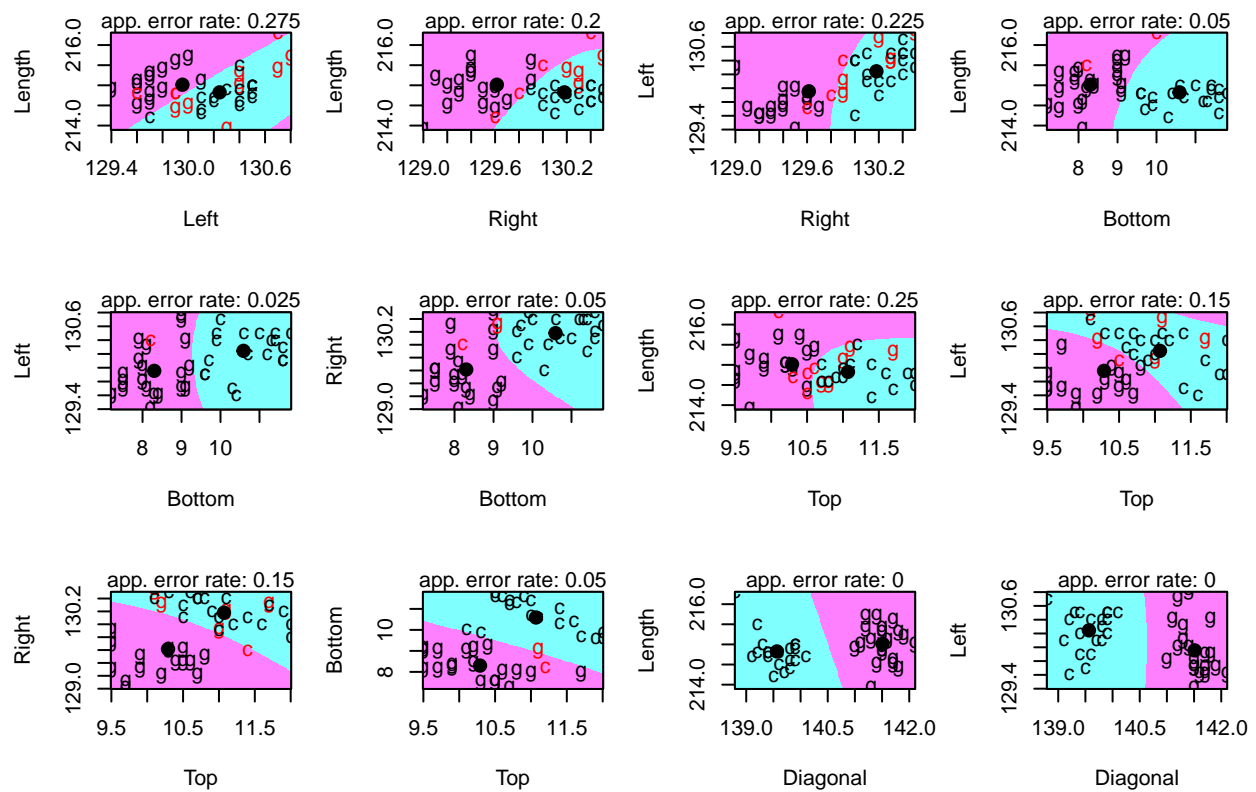
	Length	Left	Right	Bottom	Top	Diagonal
counterfeit	214.8225	130.3137	130.1962	10.51500	11.14875	139.4212
genuine	214.9575	129.9400	129.7463	8.30625	10.13750	141.5188

5.6 Evaluación del error: Accuracy Table

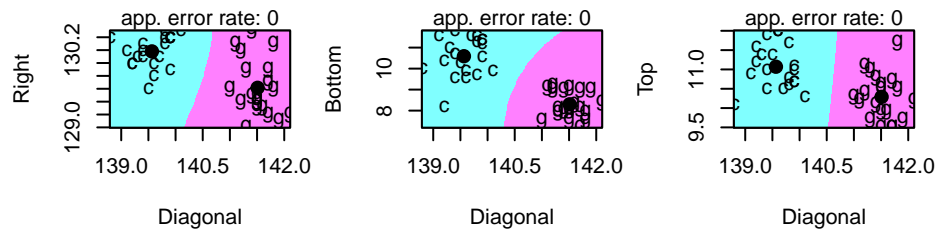
	Clase predicha	
Clase real	counterfeit	genuine
counterfeit	20	0
genuine	0	20

```
[1] "test_error = 0 %"
```

5.7 Visualización de las clasificaciones



Partition Plot



6 Analisis Discriminante Cuadrático Robusto (RQDA)

7 Máquinas de Soporte Vectorial

7.1 Datos

7.2 Grafico datos

7.3 Búsqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel lineal

7.3.1 Mejor modelo según hiperparametro

7.4 Predicciones del Modelo

7.5 Búsqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel polynomial

7.5.1 Mejor modelo según hiperparametro

7.6 Predicciones del Modelo

7.7 Búsqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel sigmoid

7.7.1 Mejor modelo según hiperparametro

7.8 Predicciones del Modelo

7.9 Búsqueda de mejor hiperparametro C (coste) y Entrenamiento del Modelo con kernel radial

7.9.1 Mejor modelo según hiperparametro

7.10 Predicciones del Modelo

8 Regresión logística (LR)

8.1 Datos

8.1.1 Boxplot todas las variables

8.1.2 Box por grupo

8.1.3 Box por variable

8.1.4 P valores

9 Clustering

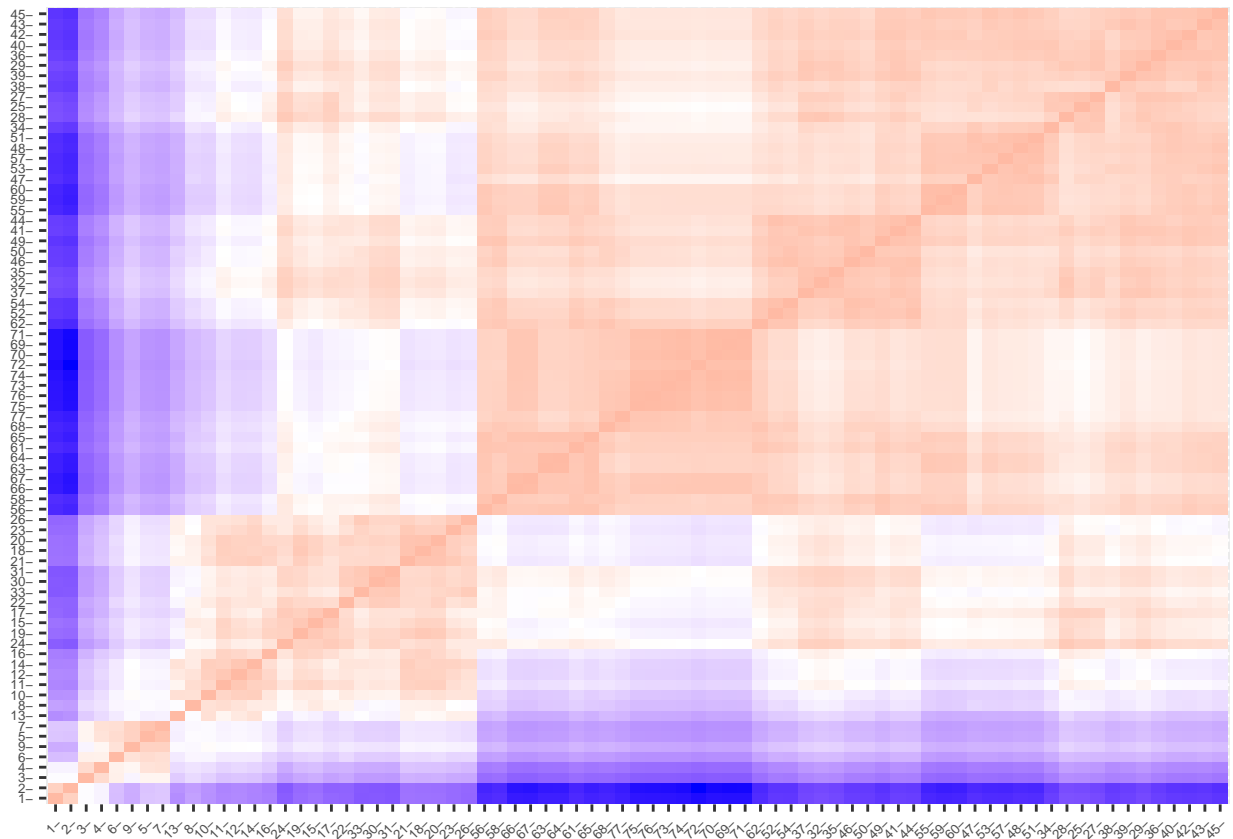
9.1 Datos

9.1.1 Se centran los datos de la matriz con omisión de datos faltantes

9.1.2 Se calculan distancias euclídeas

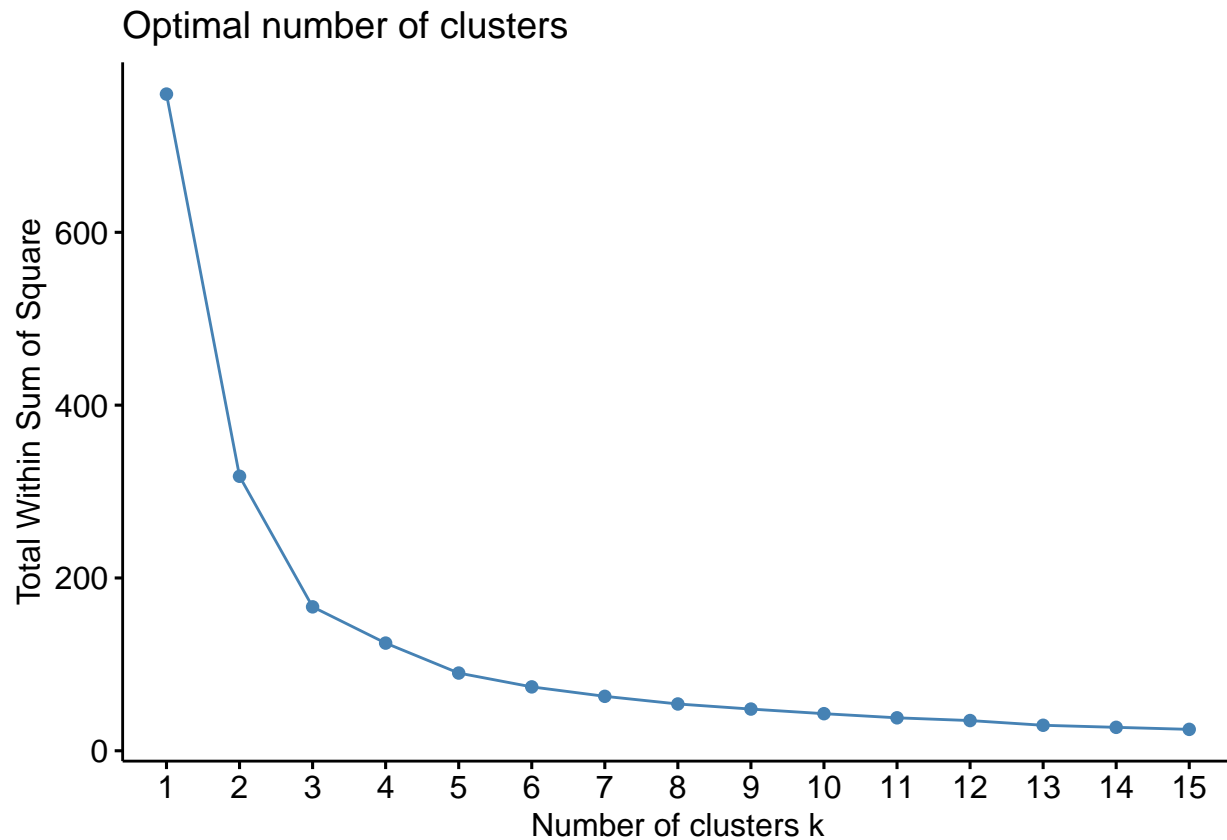
	1	2	3	4	5
1	0.00	1.20	3.62	3.98	5.89
2	1.20	0.00	3.62	3.98	5.96
3	3.62	3.62	0.00	1.67	2.98
4	3.98	3.98	1.67	0.00	2.05
5	5.89	5.96	2.98	2.05	0.00

9.2 Heatmap de la matriz de distancias entre observaciones del dataset



9.3 Cluster No Jerarquico: K-means (x centroides)

9.3.1 Selecccion de k en base al método elbow.



9.3.2 Resultado

K-means clustering with 4 clusters of sizes 2, 49, 20, 6

Cluster means:

	Mundiales	Puntos	PJ	PG	PE	PP	Rendimiento
1	2.7675086	4.2691286	3.6005910	4.3908656	2.6855720	1.6001503	2.0559176
2	-0.6429335	-0.5030788	-0.5884818	-0.4854271	-0.5722902	-0.6546405	-0.4081216
3	0.7350695	0.2728720	0.5181639	0.2365380	0.5243563	0.9717633	0.4272586
4	1.8778888	1.7758610	1.8785244	1.7122397	2.0306582	1.5736360	1.2234919
	GF	GC	Campeón				
1	4.3330839	3.0351188	2.9177024				
2	-0.5247137	-0.6377553	-0.3382843				
3	0.3653171	0.8078409	-0.3382843				
4	1.6230770	1.5038260	2.9177024				

Clustering vector:

```
[1] 1 1 4 4 4 4 3 4 3 3 3 3 3 3 3 3 3 3 3 3 2 3 2 2 2 3 3 2 3 2 2 2 2 2
[39] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[77] 2
```

Within cluster sum of squares by cluster:

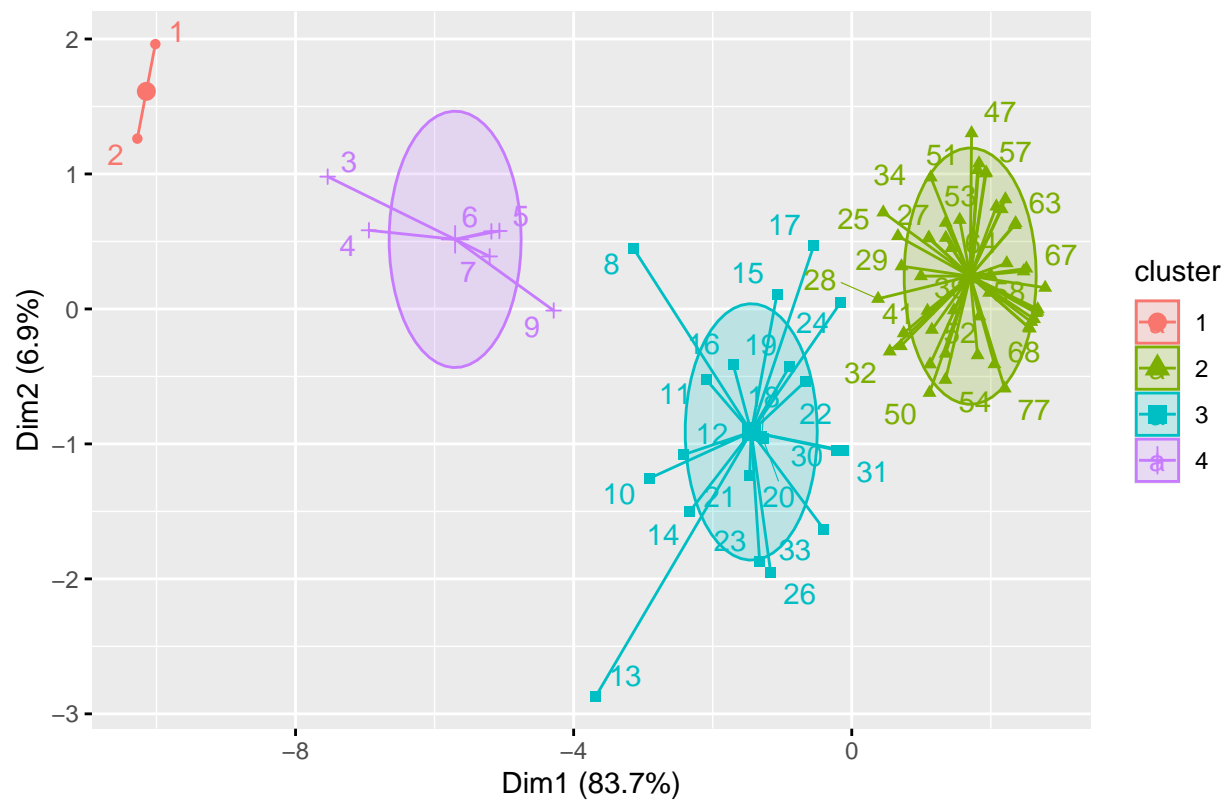
```
[1] 0.7208673 66.8240561 43.4838941 13.5951762
```

(between_SS / total_SS = 83.6 %)

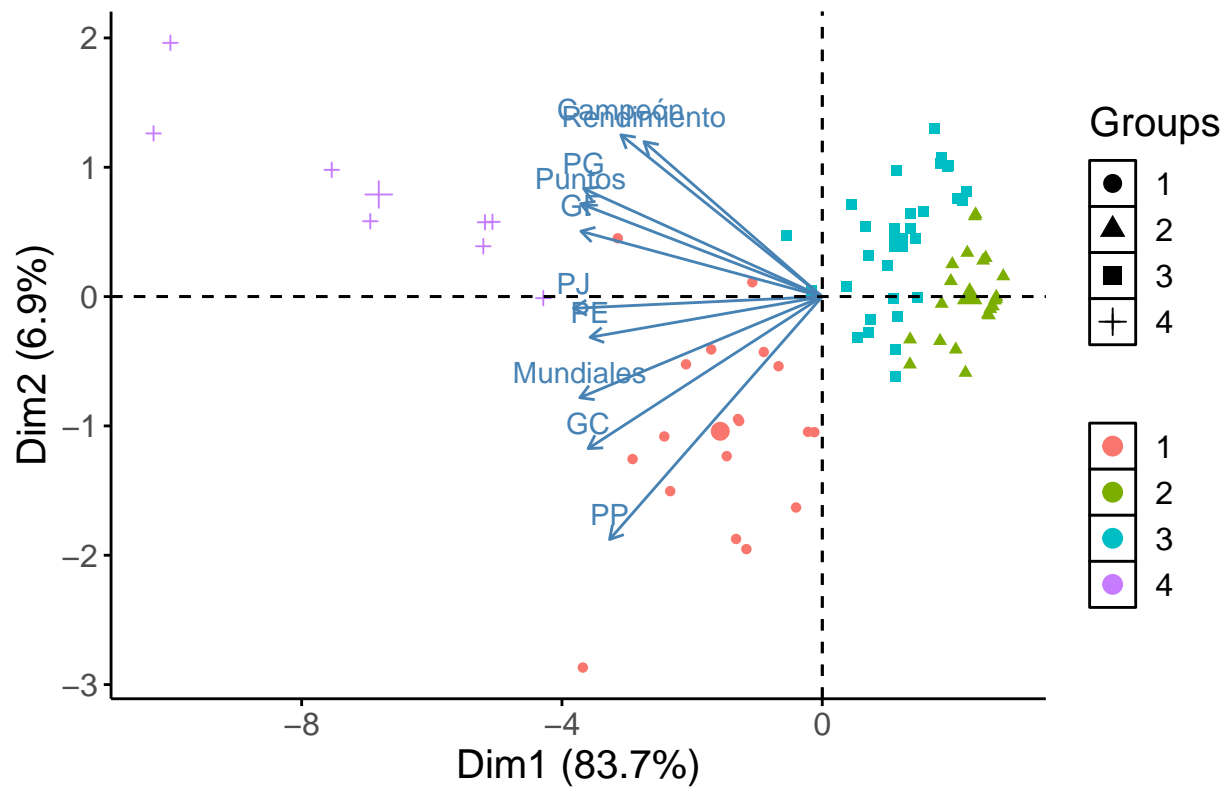
Available components:

[1]	"cluster"	"centers"	"totss"	"withinss"	"tot.withinss"
[6]	"betweenss"	"size"	"iter"	"ifault"	

K-means con k = 4

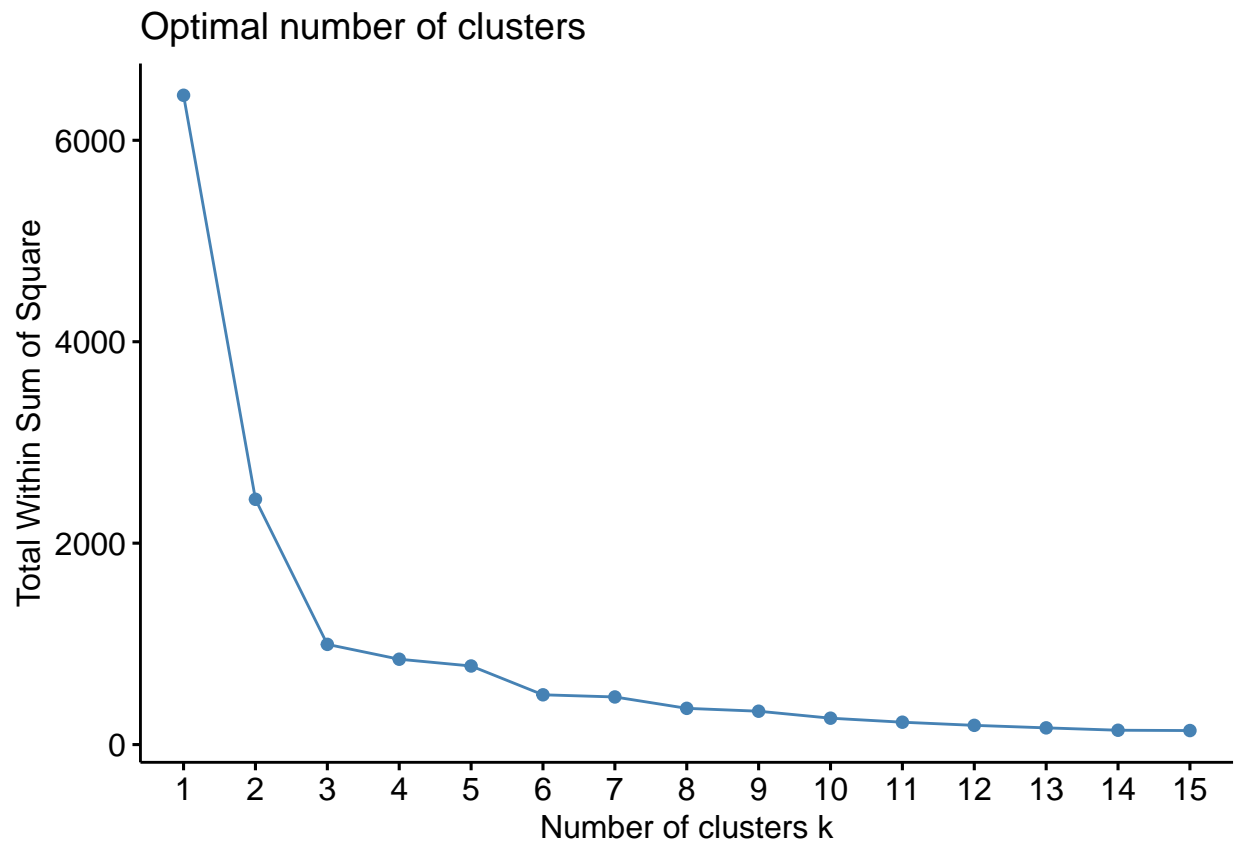


9.3.3 Grafico Cluster con PCA



9.4 Cluster No Jerarquico: K-medoids clustering (con centro en observación más representativa)

9.4.1 Selección de k con distancia de Manhattan como medida de similitud



Medoids:

	ID	Mundiales	Puntos	PJ	PG	PE	PP
[1,]	7	1.7410242	1.4614282	1.6119284	1.4419197	1.4084900	1.6796929
[2,]	20	0.7145398	0.2184936	0.4879017	0.1886176	0.4261193	1.0433517
[3,]	41	-0.3119446	-0.4029737	-0.3767342	-0.4011715	-0.3597773	-0.2293308
[4,]	66	-0.9278352	-0.6471216	-0.8090522	-0.6223425	-0.7527256	-1.0247573

	Rendimiento	GF	GC	Campeón
[1,]	1.1330108	1.6984596	1.5101018	2.9177024
[2,]	0.3729700	0.2058828	0.6816975	-0.3382843
[3,]	-0.1156277	-0.4047168	-0.4479447	-0.3382843
[4,]	-1.2014004	-0.6534796	-0.9374563	-0.3382843

Clustering vector:

```
[1] 1 1 1 1 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 2 3 3 2 2 3 2 3 3 3 3
[39] 3 3 3 3 3 3 3 3 4 4 3 3 4 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
[77] 4
```

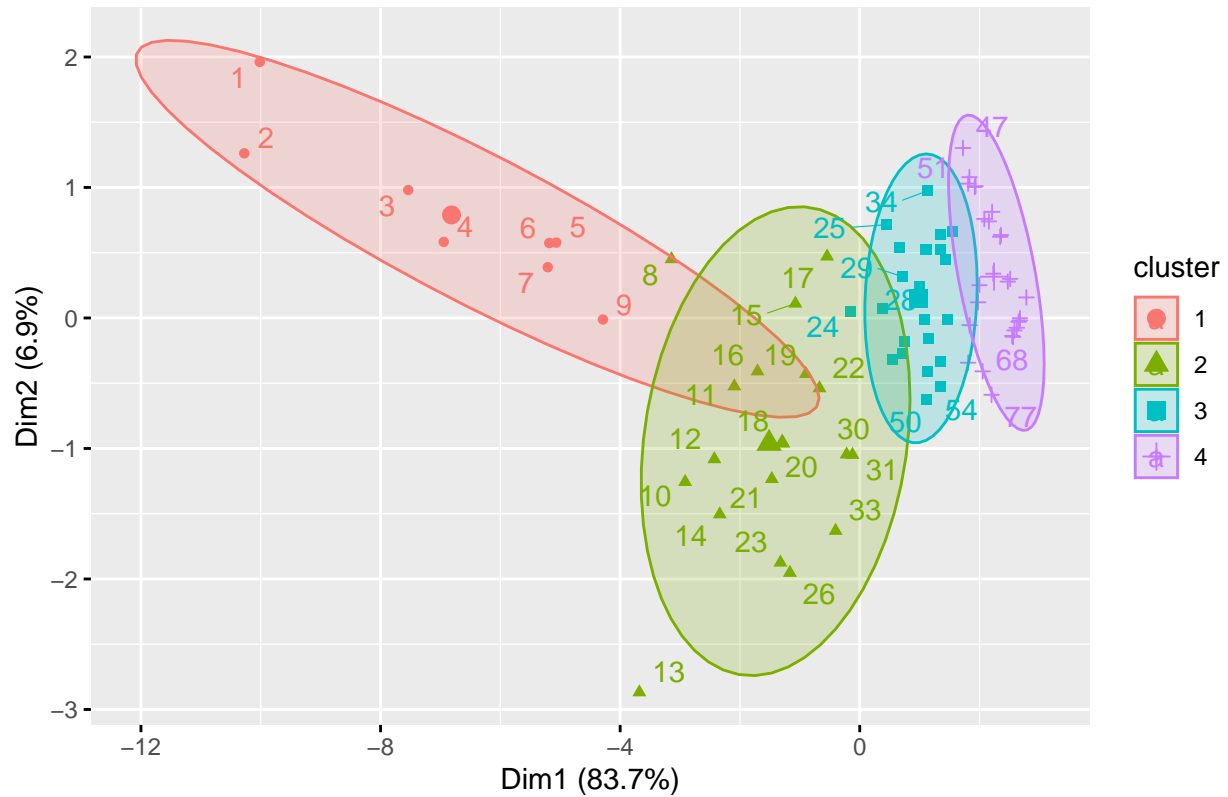
Objective function:

```
build swap
2.685052 2.573222
```

Available components:

[1]	"medoids"	"id.med"	"clustering"	"objective"	"isolation"
[6]	"clusinfo"	"silinfo"	"diss"	"call"	"data"

Resultados clustering 'Partitioning Around Medoids' con $k = 4$



9.5 Cluster Jerárquicos

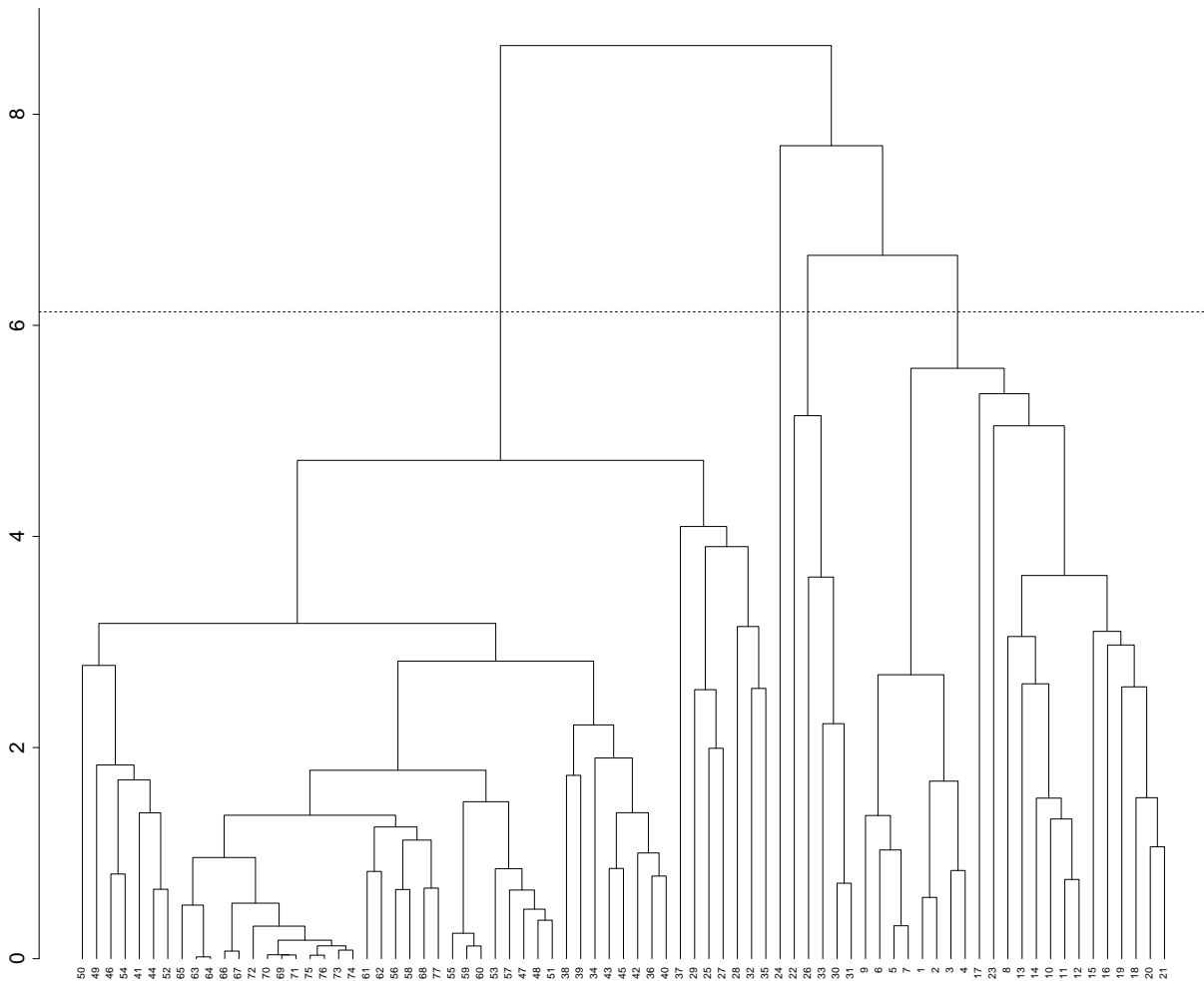
9.6 Modelo óptimo considerando distintas matrices de distancias y linkage intercluster

Table 1: Tabla de los distintos modelos -considerando distintas matrices de distancias y linkage intercluster- y sus respectivos coeficientes cofeneticos (orden descendente)

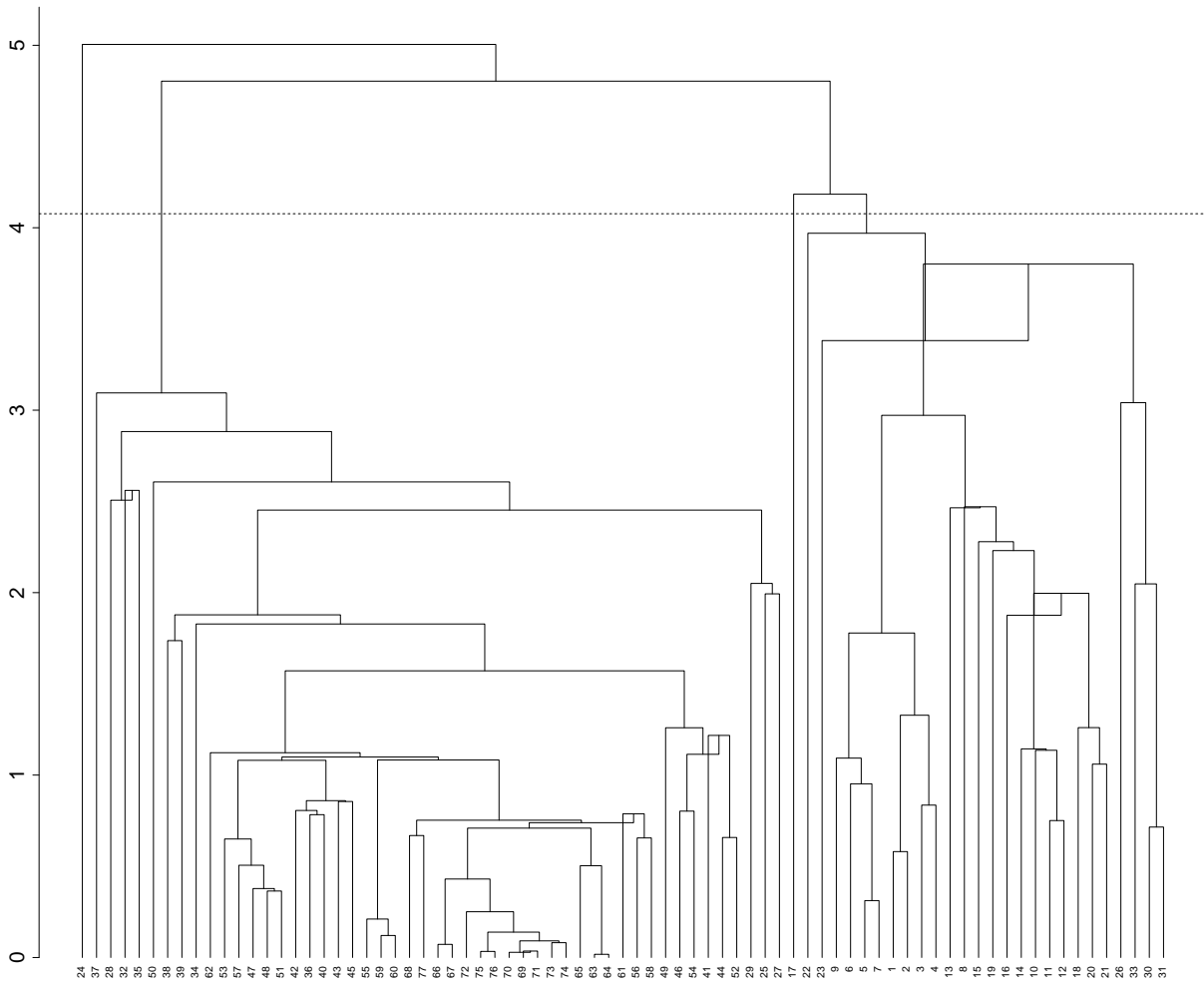
distancias	metodos linkage	coeficiente cophenetic
canberra	average	0.9529749
canberra	centroid	0.9480814
canberra	single	0.9422822
canberra	complete	0.9140296
manhattan	single	0.9042467
euclidean	centroid	0.8953428
minkowski	centroid	0.8953428
maximum	centroid	0.8822692
canberra	ward	0.8809276
manhattan	average	0.8751172
euclidean	average	0.8736778
minkowski	average	0.8736778
maximum	average	0.8664584
manhattan	centroid	0.8647472
euclidean	single	0.8634253

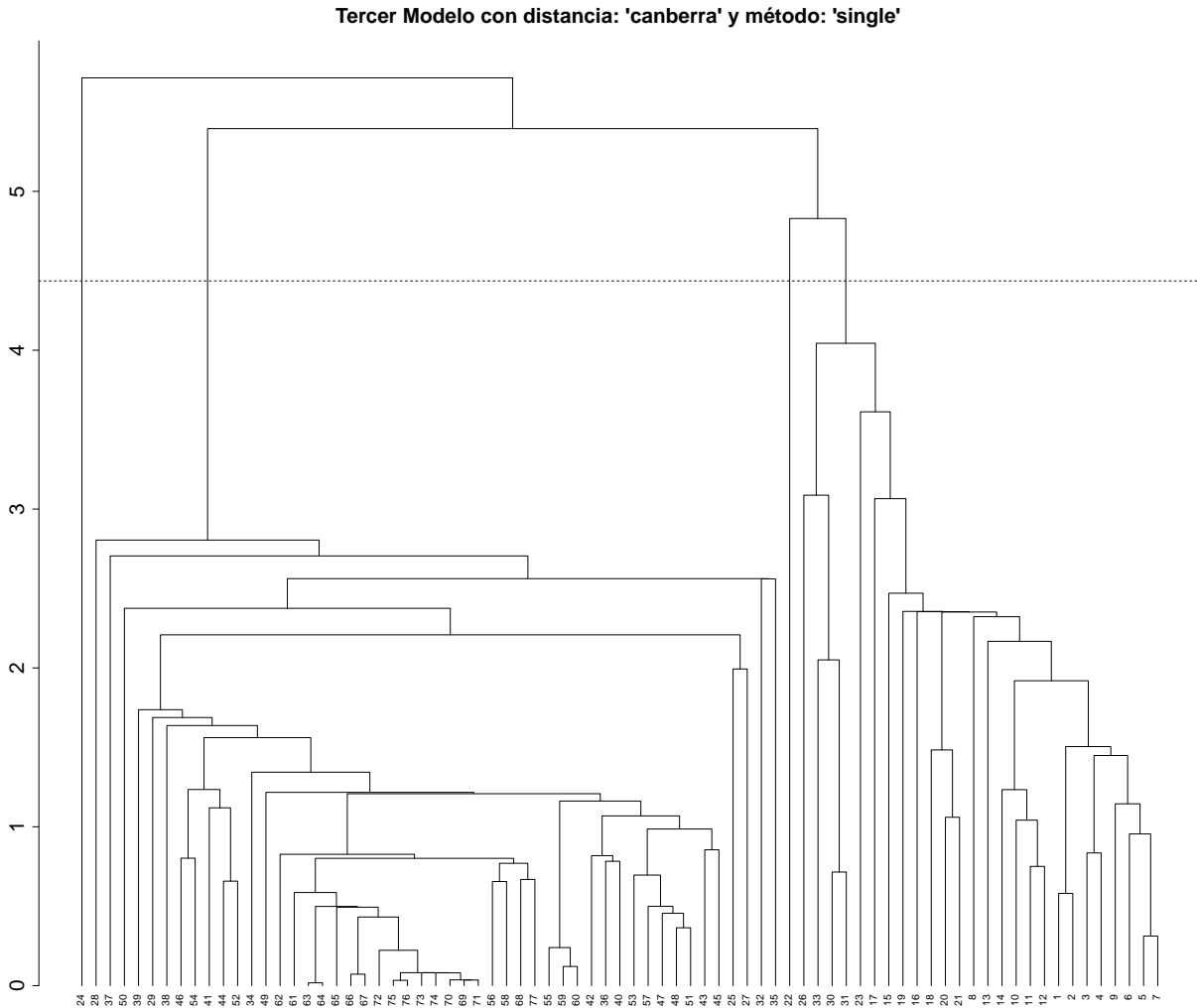
minkowski	single	0.8634253
euclidean	complete	0.8568187
minkowski	complete	0.8568187
manhattan	complete	0.8251128
maximum	single	0.8023554
maximum	ward	0.7302932
maximum	complete	0.7076196
manhattan	ward	0.6194762
euclidean	ward	0.6178955
minkowski	ward	0.6178955
binary	complete	NA
binary	average	NA
binary	single	NA
binary	centroid	NA
binary	ward	NA

Modelo Óptimo con distancia: 'canberra' y método: 'average'



Segundo Modelo con distancia: 'canberra' y método: 'centroid'





9.7 Estudio de la tendencia de clustering

[1] 0.8783879

[1] "Los datos presentan agrupamientos importante, con el estadístico Hopkins ≤ 0.75 "

10 PCA

10.1 Exploracion de datos

10.2 Analisis PCA

10.3 Varianza explicada por las CP

desviación estándar de cada componente principal

varianza

10.4 Representaciones Graficas: BIPLLOT

10.5 Proporción de Varianza Explicada

10.6 Proporción acumulada en las componentes

10.7 FactomineR

10.8 Representacion

10.9 Representacion de Variables

Grafico muestra: - %varianza explicada por cada componente (dim1 y dim2)

10.10 Contribuciones de las variables en los ejes

las contribuciones a la inercia de los ejes sirven para detectar los variables mas relevantes en cada.

10.11 Conclusión

- Considerando el criterio de Kaiser (más el ajuste derivado de las simulaciones de Montecarlo) que admite autovalores hasta 0.7, las componentes que se emplearían son: .
- Considerando el criterio de variabilidad explicada del componente, eligiendo aquellas cuyo % de variabilidad no sea menor a 5%, las componentes que se emplearían son: .

11 Sesión

R version 4.2.0 (2022-04-22)

Platform: x86_64-pc-linux-gnu (64-bit)

Running under: Ubuntu 20.04.4 LTS

Matrix products: default

BLAS: /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3

LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/liblapack.so.3

locale:

```
[1] LC_CTYPE=es_AR.UTF-8      LC_NUMERIC=C
[3] LC_TIME=es_AR.UTF-8      LC_COLLATE=es_AR.UTF-8
[5] LC_MONETARY=es_AR.UTF-8  LC_MESSAGES=es_AR.UTF-8
[7] LC_PAPER=es_AR.UTF-8     LC_NAME=C
[9] LC_ADDRESS=C             LC_TELEPHONE=C
[11] LC_MEASUREMENT=es_AR.UTF-8 LC_IDENTIFICATION=C
```

attached base packages:

```
[1] grid      stats      graphics  grDevices  utils      datasets  methods
[8] base
```

other attached packages:

```
[1] mclust_5.4.9      dplyr_1.0.9      raster_3.5-15    sp_1.4-7
[5] cluster_2.1.3     readxl_1.4.0     caret_6.0-92     lattice_0.20-45
[9] e1071_1.7-9       biotools_4.2     klaR_1.7-0       MVN_5.9
[13] reshape2_1.4.4    nortest_1.0-4    gplots_3.1.3     vcd_1.4-9
[17] factoextra_1.0.7  FactoMineR_2.4   broom_0.8.0      MASS_7.3-56
[21] htmltools_0.5.2   moments_0.14.1   corrplot_0.92    skimr_2.1.4
[25] jsonlite_1.8.0    formattable_0.2.1 tibbletime_0.1.6  readr_2.1.2
[29] ggthemes_4.2.4    rlang_1.0.2      gghighlight_0.3.2 scales_1.2.0
[33] lubridate_1.8.0   colorRamps_2.3.1 RColorBrewer_1.1-3 ggbeeswarm_0.6.0
[37] ggplot2_3.3.6     tibble_3.1.7     tidyr_1.2.0      kableExtra_1.3.4
[41] janitor_2.1.0     stringr_1.4.0    knitr_1.39
```

loaded via a namespace (and not attached):

```
[1] utf8_1.2.2        questionr_0.7.7   tidyselect_1.1.2
[4] htmlwidgets_1.5.4 combinat_0.0-8    pROC_1.18.0
[7] munsell_0.5.0     codetools_0.2-18 DT_0.22
[10] future_1.25.0     miniUI_0.1.1.1    withr_2.5.0
[13] colorspace_2.0-3  energy_1.7-10     highr_0.9
[16] rstudioapi_0.13   leaps_3.1         stats4_4.2.0
[19] ggsignif_0.6.3    listenv_0.8.0     labeling_0.4.2
[22] repr_1.1.4        mnormt_2.0.2     bit64_4.0.5
[25] farver_2.1.0      parallelly_1.31.1 vctrs_0.4.1
[28] generics_0.1.2    ipred_0.9-12      xfun_0.30
[31] R6_2.5.1          bitops_1.0-7      reshape_0.8.9
[34] promises_1.2.0.1  vroom_1.5.7       nnet_7.3-17
[37] beeswarm_0.4.0    gtable_0.3.0      globals_0.15.0
[40] timeDate_3043.102 systemfonts_1.0.4 scatterplot3d_0.3-41
[43] splines_4.2.0     rstatix_0.7.0     ModelMetrics_1.2.2.2
[46] yaml_2.3.5        abind_1.4-5       backports_1.4.1
[49] httpuv_1.6.5      tools_4.2.0       lava_1.6.10
[52] psych_2.2.3       ellipsis_0.3.2    proxy_0.4-26
```

[55] Rcpp_1.0.8.3	plyr_1.8.7	base64enc_0.1-3
[58] purrr_0.3.4	ggpubr_0.4.0	rpart_4.1.16
[61] zoo_1.8-10	haven_2.5.0	ggrepel_0.9.1
[64] magrittr_2.0.3	data.table_1.14.2	lmtest_0.9-40
[67] tmvnsim_1.0-2	gsl_2.1-7.1	hms_1.1.1
[70] mime_0.12	evaluate_0.15	xtable_1.8-4
[73] compiler_4.2.0	KernSmooth_2.23-20	crayon_1.5.1
[76] later_1.3.0	tzdb_0.3.0	boot_1.3-28
[79] Matrix_1.4-1	car_3.0-13	cli_3.3.0
[82] parallel_4.2.0	gower_1.0.0	forcats_0.5.1
[85] pkgconfig_2.0.3	flashClust_1.01-2	terra_1.5-21
[88] recipes_0.2.0	xml2_1.3.3	foreach_1.5.2
[91] svglite_2.1.0	vipor_0.4.5	hardhat_0.2.0
[94] webshot_0.5.3	prodlim_2019.11.13	rvest_1.0.2
[97] snakecase_0.11.0	digest_0.6.29	rmarkdown_2.14
[100] cellranger_1.1.0	shiny_1.7.1	gtools_3.9.2
[103] lifecycle_1.0.1	nlme_3.1-157	carData_3.0-5
[106] viridisLite_0.4.0	fansi_1.0.3	labelled_2.9.1
[109] pillar_1.7.0	GGally_2.1.2	fastmap_1.1.0
[112] httr_1.4.3	survival_3.3-1	glue_1.6.2
[115] iterators_1.0.14	bit_4.0.4	class_7.3-20
[118] stringi_1.7.6	caTools_1.18.2	future.apply_1.9.0